Comprehensive Public Report: From Digital Commons to Metadata Silo

Subject: A Comparative Investigation into the Decommissioning of the Digital Library of India (DLI), the Structural Limitations of the National Digital Library of India (NDLI), and the Public Commons Crisis in Indian Digital Archiving
Target Audience: Citizens, Independent Researchers, Scholars, and Public Knowledge Advocates

SECTION 1: Executive Summary & Historical Context

The democratic promise of the internet lies in universal, uninhibited access to human knowledge. In early post-2000 India, this ideal began materializing through the **Digital Library of India (DLI)**, an ambitious, state-supported digital repository scanning and preserving out-of-print, historical, religious, philosophical, and vernacular texts without paywalls or algorithmic restriction.
Over the last decade, however, the digital infrastructure of public knowledge in India has undergone a quiet yet catastrophic regression:

1. **The Decommissioning of DLI:** The original DLI servers (dli.ernet.in) were taken offline around 2017. The public was not provided with an official, open-access mirror of its 550,000+ digitized books and rare manuscripts.
2. **The Rescue by Civil Society:** The survival of these invaluable historical works was not secured by state institutions, but by open-access activist Carl Malamud and Public.Resource.Org, who preserved and mirrored the raw collection onto the Internet Archive (Public Library of India).
3. **The Rise of NDLI (A Metadata Aggregator):** In place of an open-access raw book repository, the state established the **National Digital Library of India (NDLI)**, hosted at IIT Kharagpur under the Ministry of Education.
4. **The Structural Breakdown:** NDLI's marketed headline metric (touting over tens of millions of "items") obscures a reality of severe academic narrowing, federated dead ends, broken search indexing, and deep catalog poverty.

When specialized topics of rich Indian heritage—such as **Jainism**—are searched on NDLI, users are often met with a handful of hits (e.g., 20–30 resources), whereas the older DLI catalog provided direct access to thousands of foundational scriptures, historical commentaries, and multilingual treatises.

SECTION 2: The Digital Library of India (DLI) — The Lost Digital Commons

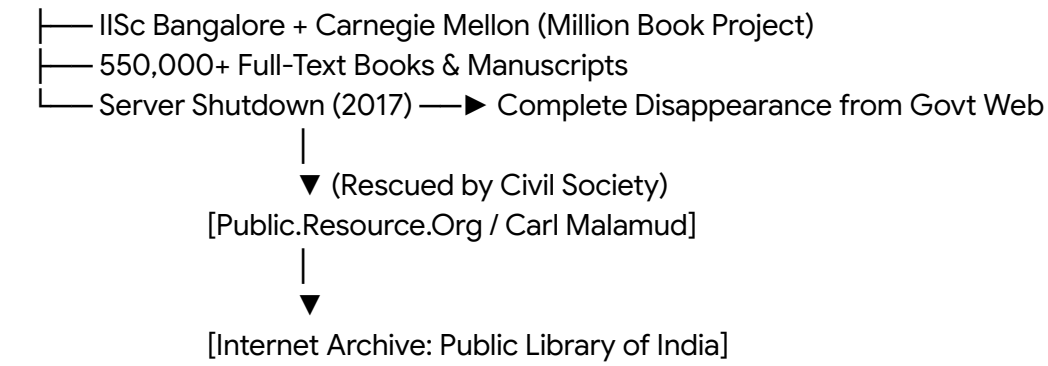
2.1 Origins and Vision

The DLI was conceptualized in the early 2000s under the leadership of the Indian Institute of Science (IISc), Bangalore, in partnership with Carnegie Mellon University's global **Million Book Project**. Its core mission was unambiguous: to digitize all significant literary, artistic, and scientific works of human history and make them freely available to the global public for preservation and discovery.

2.2 Architectural and Curation Model

- **Direct File Repository:** The DLI did not operate as a middleman or metadata pointer. It stored raw, digitized file scans (TIFF, DjVu, and PDF formats) directly on reachable host servers.
- **Unrestricted Public Access:** There were no user tiers, subscription gates, or proprietary publisher restrictions. Anyone with an internet connection could search the catalog and download full-text works directly.
- **Broad and Multilingual Scope:** The collection spanned more than 50 Indian languages, capturing diverse cultural, religious, and philosophical traditions (including large collections on Jainism, Buddhism, Vedanta, Islamic scholarship, regional folktales, and British-era administrative records).

SECTION 3: The Decommissioning of DLI and the Civil Society Rescue

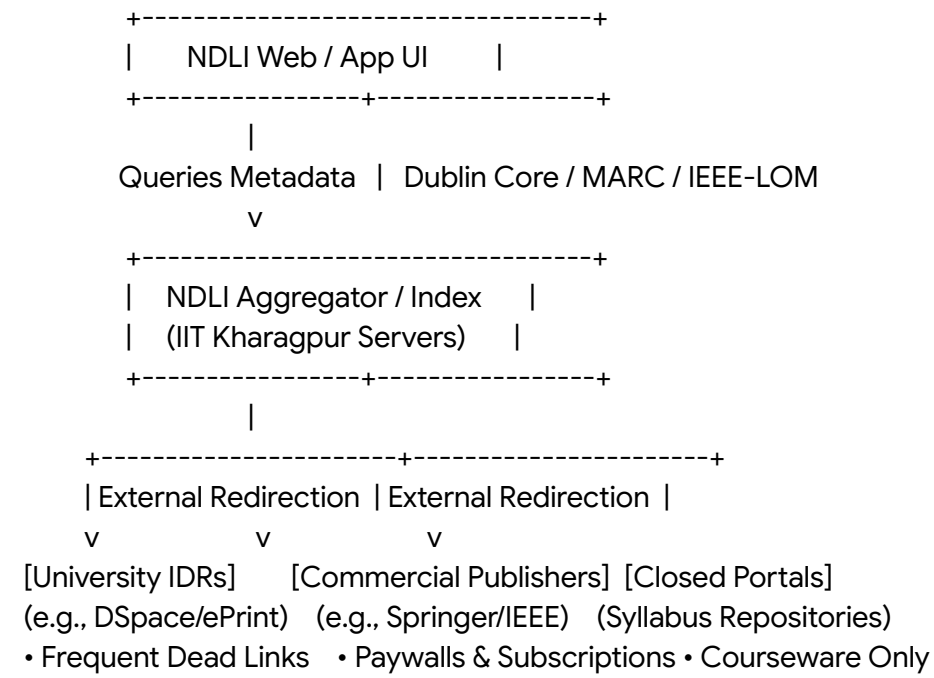

[DLI Phase (2000s - 2017)]
├── IISc Bangalore + Carnegie Mellon (Million Book Project)
├── 550,000+ Full-Text Books & Manuscripts
└── Server Shutdown (2017) ──► Complete Disappearance from Govt Web
▼ (Rescued by Civil Society)
[Public.Resource.Org / Carl Malamud]
▼
[Internet Archive: Public Library of India]

When the DLI portal was retired without an integrated, direct-download public successor, a critical segment of India's digitized cultural heritage was nearly erased from the open web. Through the intervention of Carl Malamud, the data was extracted from sunsetting servers and organized into the **Public Library of India** collection on the Internet Archive. Without this civil society intervention, hundreds of thousands of rare Indian texts would have been lost to researchers, students, and citizens.

SECTION 4: Architecture of the National Digital

Library of India (NDLI)

To understand why NDLI fails as an exploratory book catalog, one must examine its structural architecture:



1. **The Middleman (Federated) Model:** NDLI does not host the vast majority of its contents. It only harvests metadata (such as author, title, and keywords).
2. **External Redirections:** When a user clicks on an item, NDLI points the browser outward to an external server (Institutional Digital Repositories, university libraries, or commercial academic publisher databases).
3. **No Direct Accountability:** As acknowledged in academic literature on NDLI, the platform assumes zero responsibility for the uptime, validity, or full-text accessibility of external host files.

SECTION 5: The "Item Count" Metric vs. True Accessibility

NDLI public communications frequently report tens of millions of resources to demonstrate scale. However, a forensic look at the composition of these items reveals a very different picture from that of a public library:

Asset Type in Aggregator	Characteristic	Utility for General Book

		Discovery
Journal Abstracts & Metadata Records	Single-page abstracts pointing to paywalled publisher portals	❌ Inaccessible without expensive subscriptions
Institutional Course Modules	Syllabus PDFs, question banks, presentation slides, and video lectures	❌ Tailored solely for formal exam coaching
Institutional Theses / Dissertations	Specialized postgraduate academic papers	⚠️ Limited to niche academic research
Complete, Out-of-Print Books	Full-text, downloadable historical/cultural books	❌ Extremely low representation compared to DLI

Counting each article abstract, video clip, and slide deck as an independent "digital resource" allows an aggregator to present large metrics while offering a tiny fraction of standalone, readable books.

SECTION 6: The "Jainism" Case Study: Catalog Poverty & Search Degradation

A clear illustration of this structural problem is seen when querying specific heritage subjects:

Case Comparison: Search Query = "Jainism"

[Digital Library of India (DLI Archive)]
├── Scope: Comprehensive manuscript & historical literature
├── Search Depth: Direct full-text keyword indexing across scans
└── Results: Thousands of rare texts, Prakrit/Sanskrit commentaries,
out-of-copyright translations, and historical analyses.

[National Digital Library of India (NDLI)]
├── Scope: Constrained institutional syllabus & external links
├── Search Depth: Surface metadata only (titles, Dublin Core keywords)
└── Results: Approx. ~24 resources (heavily fragmented, lesson notes,

broken redirections, or narrow journal entries).

Why This Discrepancy Occurs:

- **Absence of Full-Text Optical Character Recognition (OCR):** DLI projects scanned and indexed raw textual layers, enabling searches *inside* the contents of books. NDLI searches primarily match surface metadata strings.
- **Exclusion of Non-Curricular Heritage Collections:** Because NDLI is optimized for competitive entrance exams (JEE, NEET, UPSC) and university syllabi, vast tracts of independent cultural, historical, and philosophical texts are excluded.
- **Aggregator Fragmentation:** Books on Jain philosophy that once existed as complete multi-volume scans are either absent or fragmented into unindexed institutional sub-collections.

SECTION 7: Systematic Comparison: DLI vs. NDLI

Parameter	Digital Library of India (DLI)	National Digital Library of India (NDLI)
Primary Institutional Custodian	IISc Bangalore / Carnegie Mellon	IIT Kharagpur / Ministry of Education
Primary Architecture	Centralized Content Repository (Raw Scans)	Federated Metadata Aggregator (Pointer Network)
Public Accessibility	100% Free, direct public downloads	Mixed; frequent paywalls, institutional logins, and session timeouts
Curatorial Philosophy	Democratic Preservation of Human Knowledge	Syllabus-Driven Academic & Exam Preparation
Link Integrity	Direct internal storage; zero external dead links	High link rot; dependent on external university servers
Preservation of Rare Manuscripts	Primary mission (550,000+ historic texts)	Secondary; incidental metadata harvesting
Current Operational	Servers dead; mirrored on	Active, heavily funded

Status	Internet Archive	national mission platform

SECTION 8: Technical Bottlenecks, Link Rot, and UI Friction

Citizens navigating the NDLI portal regularly encounter technical obstacles stemming from its core design:

1. **The "Middleman Latency" Problem:** Every search query triggers federated lookups across external databases. This leads to slow page loading, connection timeouts, and broken search sessions.
2. **Link Rot & Dead Ends:** Academic research confirms that NDLI takes no custodial responsibility for content delivery. When partner universities change URL paths or decommission local servers, the NDLI record remains an empty shell.
3. **Format Chaos:** While DLI standardized on readable full scans, NDLI returns a disjointed mix of .pdf, .html, .doc, video streams, and proprietary viewer plugins, creating severe viewing friction on desktop and mobile browsers.
4. **Subscription Walls:** Thousands of search results route users directly to commercial academic publisher paywalls (such as Springer or IEEE), defeating the purpose of an open national public library.

SECTION 9: Socio-Political & Epistemic Implications

The shift from an open digital repository to a curated, syllabus-focused metadata network carries broader consequences for researchers and the reading public:

- **Narrowing of Intellectual Discovery:** A true library encourages serendipitous discovery across unexpected historical, philosophical, and critical texts. An aggregator optimized for standard exams narrows intellectual discovery to pre-approved syllabi.
- **Institutional Gatekeeping:** Knowledge access is once again conditioned on institutional affiliations, university email credentials, and commercial journal subscriptions.
- **Cultural and Historical Amnesia:** When public bodies allow direct-scan repositories to disappear from official servers, access to local histories, regional-language texts, and minority philosophical traditions (such as classical Jain or Buddhist literature) is significantly reduced for ordinary citizens.

SECTION 10: Citizen Action Guide & Policy Recommendations

What Citizens and Researchers Can Do Today

- **Access the Rescued DLI Collection:** Bypass aggregator dead ends by searching the preserved collection directly on the Internet Archive: Public Library of India.

- **Utilize Open-Access Scholarly Repositories:** For independent access to open Indian academic research, use Shodhganga (the national digital repository of Indian doctoral theses) directly, rather than through NDLI's search wrapper.

Policy Reforms Needed for a True National Digital Commons

1. **Restore Full-Text Hosting:** The Ministry of Education and IIT Kharagpur must transition NDLI from a metadata aggregator to a true digital repository that directly stores and serves out-of-copyright and public-domain works.
2. **Re-integrate the Full DLI Corpus:** The government should officially ingest and host the entire 550,000+ DLI historical collection on permanent, high-speed, direct-download public servers.
3. **Mandate Full-Text OCR & Multilingual Search:** Modernize search indexes to match against the full text of books across all classical and regional Indian languages, rather than relying exclusively on Dublin Core metadata.
4. **Enforce Public Commons Accessibility:** Remove all external commercial paywalls and session restrictions from materials funded through public resources.